DOCUMENT RESUME

ED 398 749 FL 024 051

AUTHOR Kuo, Pinmin
TITLE The Correlation of Discourse Markers and Discourse Structure.
PUB DATE 94
NOTE 24p.; In: Pragmatics and Language Learning. Monograph Series, Volume 5, p208-30, 1994; see FL 014 038.
PUB TYPE Reports - Evaluative/Feasibility (142)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS Discourse Analysis; *Grammar; *Language Patterns; Language Research; Linguistic Theory; *Mandarin Chinese; *Pragmatics; *Research Methodology; Structural Analysis (Linguistics); Written Language
IDENTIFIERS *Connectives (Grammar); *Discourse Markers

ABSTRACT

In discourse analysis, connectives have been widely suggested as linguistic markers to indicate the logical linkage between utterances. However, the understanding of the interactions among various kinds of connectives in discourse has been limited. A method of quantifying the overall correlation between different kinds of connectives occurring on coherent texts is proposed. This analysis of discourse structures focuses on two written texts in Mandarin Chinese, and both illustrates the complexity of interactions among various connectives and reveals patterns of connectives indicating the logic structure in discourse. Methodology used in coding and quantifying the Chinese connectives within sentences and paragraphs and data summaries are presented. Theoretical and pedagogical implications are considered. It is concluded that this methodology, a numerical measurement of correlation coefficients, can be used effectively for: (1) showing that the complex sentence in Mandarin represents a topic continuity; (2) helping to prepare language textbook content; (3) confirming a taxonomy of coherent relationships; and (4) helping to generalize the modification direction for the inferential function denoted by each connective group. Contains 30 references. (MSE)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

Pragmatics and Language Learning
Monograph Series, Volume 5, 1994

THE CORRELATION OF DISCOURSE MARKERS AND DISCOURSE STRUCTURE

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

Pinmin Kuo
University of Illinois

PERMISSION TO REPRODUCE AND DISSEMINATE THIS MATERIAL HAS BEEN GRANTED BY
TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

ED 398 749

ABSTRACT

In discourse analysis, connectives have been widely suggested as linguistic markers to indicate the logic linkage between utterances. However, the understanding of the interactions among various kinds of connectives in discourse was limited. The overall pictures of discourse structures, thus, remain unclear.

The purpose of this paper is to propose a method to quantify the overall correlation between different kinds of connectives occurring in coherent texts. My survey of discourse structures is focused on the written text in Mandarin Chinese. Based on this quantitative study, the complexity of the interaction among various kinds of connectives is illustrated. Furthermore, the patterns of connectives which indicate the logic structure in discourse are also revealed.

Recently the correlation method was applied to linguistic elements for measurements of relatedness in dialect affinity. In this study, the numerical measurement of correlation coefficients is used to help us interpret the relations of connectives in coherent texts. Based on the thorough measurement, in my view, a better understanding of the variety of discourse structures can be reached.

1. SCOPE OF THE STUDY

Discourse connectives are regarded as the main linguistic device available for the writer to guide the reader's inferences about the text. Conversely, the reader's interpretation of the logical flow of the discourse is largely based on the distribution of discourse connectives. The logical linkage of a discourse, like the skeleton of a human body, can be illustrated by the use of discourse connectives. Thus, my primary concern in this research is to explore the relationship between discourse connectives and patterns of inference in a coherent plan in order to establish the discourse structure of a text. This study explores the relationship between the contribution of connectives to a higher level of discourse structure.

In order to investigate the overall construction of a discourse, the use of connectives must be investigated. First, one has to consider questions such as, what is a discourse connective?

FL 024051

BEST COPY AVAILABLE

What is discourse structure? And how does the interaction between connectives reflect the writer's plan and help the reader interpret a fragment of text?

Examples (1) and (2) illustrate some points that will be focused on in the study of discourse connectives. First, in a sentence-based linguistic theory, connectives are known to be used for connecting clauses, phrases, and words. In (1), *keshi* 'but', in the second clause connects two clauses within a sentence: the clause introduced by *keshi* and its preceding clause. However, this analysis is not able to explain *keshi* in (2). On the one hand, *keshi* in clause 4 introduces a new sentence; no clause precedes *keshi* in this sentence. On the other hand, simply connecting the clause preceding of *keshi* (clause 3) and the *keshi*-introduced-clause (clause 4) does not help the reader interpret the whole discourse. Intuitively, in this case, rather than two clauses, larger units of discourse are connected by *keshi*. How large is the scope, then, if *keshi* is used to connect more than two clauses? There must be some general principles of the use of *keshi* that the reader can follow to interpret the discourse. Without knowing the macroclausal (or macrosyntactic) and the clausal (or syntactic)[1] uses of *keshi*, the reader would not know which utterances are connected by *it*.

(1) 1. Ta yiwei ziji shi tie zuo de
he think himself be iron make Nom[2]
'He thought that he was made of iron,'

2. **keshi** ganqing ta ye hui bing.
but actually he too will sick
'but actually he too could be sick.' (*Luotuo Xiangzi* p.11)[3]

(2) 1. Ta **hai** qiang da zhe jingshen,
he still force P energy
'He was forcing his energy'

2. **buzhuan wei** hun yi tian de jiaogu,
not-only because make one day Nom food
'not only because he need to work to fill his stomach for the day,'

3. **erqie** yao jixu zhe jichu mai che de qian.
but-also want continue P save buy rickshaw Nom money
'but also he had to continue saving his money to buy the rickshaw.'

4. **Keshi** qiang da zhe jingshen yongyuan bushi jian tuodang de shi:
but force keep P energy always not piece proper Nom thing
'But forcing your energy is never a good thing to do:'

5. la qi che **lai**,
pull P rickshaw when
'when he was pulling a rickshaw'

6. ta bu neng zhuanxin yizhi de pao,
he not able concentrate Nom run
'he could not keep his mind on the job and run straight along,'

7. **haoxiang** lao xiang zhe xie shenme,
like always think P some what
'it was as if he was always thinking of something,'

8. **yue** xiang
the-more think
'and the more he thought'

9. **bian yue** haipa
then the-more afraid
'the more afraid

10. **yue** qibuping.
the-more upset
'and upset he became.' (*Luotuo Xiangzi* p.10)

Like *keshi*, many other connectives function macroclausally in a coherent discourse. As such, the significance of the function played by connectives can be accounted for only in a discourse-based analysis.

In addition to the function of each single connective in discourse, the second point that will be focused on in this research is the interaction between connectives. For instance, in (2), in addition to the use of *keshi* 'but' in clause 4, other connectives are used to serve different transition functions in the discourse (highlighted in boldface): *Hai* 'still, or again' is used in clause 1; *buzhuan* 'not only', and *wei* 'because' are used in clause 2; *erqie* 'but also' is used in clause 3. In clause 5, lai 'at...circumstance' is used; in clause 7, *haoxiang* 'as if' is used; and in clauses 8, 9 and 10, *yue* 'the more...the more' and bian 'then' are used. The interaction of connectives will also be useful to interpret the logical linkage in a larger scope of discourse. For instance, knowing that the connectives *buzhuan* 'not only' and *erqie* 'but also' are used mostly as a pair will help the reader understand that clauses 3 and 4 are closely congruent as a larger statement serving an elaboration function in the discourse.

After knowing the feature of each connective and the interaction between connectives, the construction of the whole discourse in terms of its logical linkages becomes explicit. The third point to be focused on in this research is the construction of the discourse based on the knowledge we obtain on the distribution of discourse connectives.

A quantitative method will be proposed to analyze the discourse connectives used in written texts in Mandarin Chinese. This quantitative study of discourse connectives investigates the interaction of discourse connectives in a communication-based discourse.

2. DATA

In this research, I limited data to the simplest type of discourse, a discourse constituted by a finite sequence of declarative and narrative statements, made by one writer. My survey of discourse connectives and the inferential relation they denoted will be focused on the written text.

The data analyzed are based on *Luotuo Xiangzi 'The Rickshaw Boy'* (1982, first printing in 1936) and *Sishi Tongtang* 'The Yellow Storm' (1983, first printing in 1946 to 1950) written by Lao She, the well-known Chinese twentieth century writer. Lao She's written language is treated as representative of modern Mandarin Chinese (Chao 1968) and is adopted as the data source in various discourse analyses. *Luotuo Xiangzi* and *Sishi Tongtang* are his famous works.

Luotuo Xiangzi in this study is based on the version published by *Sichuan Renmin*

Chubanshe (1982). I transcribed this story into the computer in Pinyin (without tonal indications). *Luotuo Xiangzi* consists of 5,126 sentences, 1,075 paragraphs in print, and a total of 149,040 characters.

The database of *Sishi Tongtang* was established by Fumiyoshi Matsumura between 1986 and 1987.[4] It consists of 27,549 sentences, and 6,201 paragraphs, in a total of 817,000 characters.

3. DISCOURSE MARKERS IN MANDARIN CHINESE

Discourse connective is not a syntactic category; rather, it is a functional term to indicate the logical linkage between utterances. In the study of discourse, although the syntactic category "connective" indeed plays an important role in terms of logical linkage, other syntactic categories such as adverbial and preposition could also play the same role. In Chinese, *guanlian ci* 'relation word' is a particular group of words which are used to connect discourse fragments. The discourse fragments can be of different scopes, such as words, phrases, clauses, sentences and paragraphs. *Guanlian ci* includes expressions in different syntactic categories and has a very similar function as a discourse connective. It has been suggested in Lü (1980:13), and *Hanyu Yufa Xiuci Cidian* (1986:171) 'A Dictionary of Chinese Grammar and Rhetoric' (edited by Dihua Zhang) that *guanlian ciyu* 'relation word/phrase' includes connectives (*lain ci*) and a particular group of adverbials (*fu ci*) and short sentences (*duan ju*) which have the function of connection.

In this study, the discourse connectives include connectives and a particular group of prepositions and adverbials which have the function of connection. Some nouns, verbs, and short sentences which may also have the "function of connection" are excluded in this study. This is primarily because there are many alternatives for the expressions conveyed by the nouns, verbs or short sentences. For instance, *tingdao zhege* 'once hearing it' functions to mark the sequence between the previous action or event and the following utterance. However, this expression is not unique in that there can be other expressions with the same pattern and the same function, such as *xiang daole zhege* 'once thinking of it', *shuodao zhe li* 'once speaking of it', *kandao zher* 'once seeing it' and so on. Other expressions of this sort are also excluded from this study, such as *mingzhidao* 'knowing', *dagaideshu ba* 'generally speaking', *duile* 'it's correct', *xiang bu dao* 'unexpected', *jiashang* 'plus, *jintian* 'today', *zuotian* 'yesterday', *mingtian* 'tomorrow.'

In consideration of the syntactic category involved, I examine the *guanlian ci* 'relation word', *lian ci* 'connective', and *guanxi ci* 'relation word' discussed in Guo (1960), Chao (1968), Lü (1980), Li & Thompson (1981), Okurowski (1986), *Hanyu Yufa Xiuci Cidian* (Zhang 1986), Li (1990), *Zhongguo Yuyanxue Da Cidian* (Chen 1989), Lee (1990), and *Xinhua Judian* (Zhang 1991) in order to give a broader view of discourse connectives in Chinese.

Based on the functions the coherence relations have in discourse, Hobbs (1979) points out that there are four requirements for a successful communication: (i) the message itself must be conveyed; (ii) the message must be related to the goals of the discourse; (iii) what is new and unpredictable in the message must be related to what the listener already knows; and (iv) the speaker must guide the listener's inference processes toward the full intended meaning

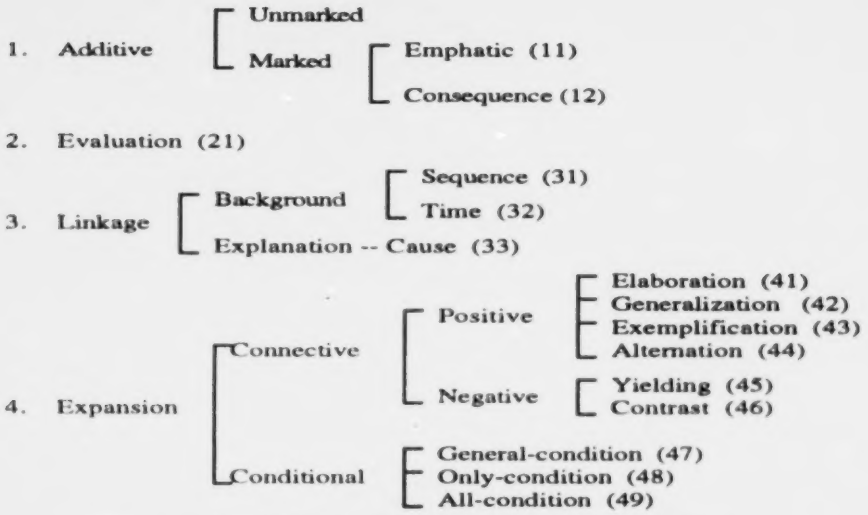


Figure 1. A Modified Taxonomy of Coherence Relations

of the message. Corresponding to each requirement is a class of coherence relations which helps the speaker satisfy the requirements. I modified the coherence relations suggested by Hobbs (1978, 1979) and provided them with a more detailed framework so that more proper divisions of inferential patterns are included. In addition, for the ease of data searching and processing, each inferential relation is given a two-digit code as shown in Figure 1. The first digit represents the upper level of the communication taxonomy, and the second digit represents the sub-group. Another task of this research will be to investigate the level of accuracy and completeness of the taxonomy specified thus far.

On the basis of the taxonomy on Figure 1 and the discourse connectives discussed in different studies, in this study Chinese discourse connectives were coded according to their uses and meanings. There are a total of 217 connectives in this study, as listed in Table 1. The first two digits of the code represent the relation group they belong to and the third and fourth digits are the sequential numbers. In the following discussion, a **connective group** will be used to represent the connectives which have the same logical relation, i.e., the first two digits of the code.

Table 1 The Coding of Connectives

1101	hai	1108	fanzheng	1201	jiusuan
1102	ye	1109	shenzhiyu	1202	er
1103	you	1109	shenzhi	1202	conger
1104	geng	1109	shenerzhiyu	1203	zhihao
1105	rengjiu	1110	zai	1204	jieguo
1106	dou	1201	jiu	1205	yizhi
1107	lian	1201	jiushi	1205	yizhiyu

1206	name	3111	disan	4101	budu
1206	na	3112	disi	4101	budan
1207	bian	3113	yibian	4101	buguang
1208	suoyi	3114	jiezhe	4101	bute
1209	yinci	3115	jiner	4101	bujin
1210	yiner	3201	congqian	4101	buzhi
1211	yushi	3201	xianqian	4101	buzhuan
1211	yushihu	3202	yiqian	4102	bingqie
1212	cai	3203	xianzai	4102	shangqie
1213	ze	3203	jinlai	4102	bing
1214	fouze	3204	tongshi	4102	erqie
1215	buran	3205	nashihou	4102	er
1216	gu	3205	dangshi	4103	jiayi
1217	erhou	3206	congci	4103	yiji
1218	yibian	3207	zicong	4105	zaishuo
2101	bucuo	3208	yihou	4106	lingwai
2102	duide	3209	ranhou	4106	ciwai
2103	guobuqiran	3210	houlai	4107	tongyang
2103	guoran	3211	weilai	4108	chule
2106	dangran	3216	qingkuang	4111	hekuang
2106	ziran	3217	zuichu	4111	erkuang
2107	shide	3218	zuihou	4112	ji
2109	zhemeyang	3219	yuanlai	4113	kuangqie
2109	zheyang	3219	yuanxian	4201	zongeryanzhi
2109	zhemezhe	3219	benlai	4201	zongzhi
2110	haozai	3219	yuanben	4202	huanjuhuashuo
2113	kongpa	3220	jizhi	4301	xiang
2115	duiyu	3220	yizhi	4302	bifang
2116	guanyu	3221	shihou	4304	fangfu
2122	yaoburan	3222	zhengdang	4305	liru
2122	buran	3222	zheng	4306	ru
2122	yaobu	3223	jieguyanr	4307	side
2126	zhengshi	3227	dangchu	4308	haosi
3101	diyi	3228	gangcai	4401	huozhe
3102	dier	3228	xianglai	4401	huoze
3103	yibian	3301	weile	4401	huo
3104	yilai	3302	jiran	4402	haishi
3105	erlai	3302	ji	4403	yi
3106	xian	3303	youyu	4404	yaome
3107	yue	3303	jianyu	4405	yuqi
3108	qici	3304	yinwei	4406	ningke
3108	suishour	3304	yin	4406	shuruo
3109	zuihou	4101	budan	4406	buru
3110	yue	4101	feidan	4406	wuning

4501	suiran	4607	buguo	4707	tang
4501	sui	4608	er	4707	tanghuo
4501	chengran	4609	rengran	4707	tangshi
4501	guran	4610	qishi	4708	wanyi
4501	zongran	4611	buliao	4709	yaoshi
4501	suishuo	4612	kexi	4709	yao
4501	suize	4613	xinger	4709	yaobushi
4502	jinguan	4614	fanzhi	4710	guoran
4503	napa	4615	xiangfan	4710	guozhen
4504	jihuo	4616	dao	4711	zhiyao
4504	jiling	4617	zhishi	4801	zhiyou
4505	jishi	4701	dehua	4802	chufei
4505	jibian	4702	jiashi	4901	bulun
4506	jiushi	4702	jiaru	4901	wulun
4602	keshi	4702	ru	4901	wulunruhe
4603	raner	4703	jiaruo	4902	buguan
4604	queshi	4703	ruruo	4903	fanshi
4604	que	4704	ruguo	4904	zong
4605	fandao	4706	ruoshi	4905	renping
4605	faner	4706	ruo	4905	ping
4606	danshi	4706	sheruo		
4606	dan	4707	tangruo		

4. METHODOLOGY AND ANALYSIS

The correlation coefficient is considered as an indicator of degree of concurrence between connectives, that is, the indicator of the closeness between every two groups. The higher the coefficient value, the closer the connective-groups are associated. Based on this concept, I calculate the correlation coefficients of all connective-groups in each topic continuity, which includes the scope of sentence and the scope of paragraph in print. The scope of sentence is recognized by the use of the full stop punctuation signs: ".", "?" and "!"; and the scope of sentence and the scope of paragraph is recognized by the indentation at thevery beginning of a discourse chunck. [5]

First, all the connectives in Table 1 are searched throughout the text of *Sishi Tongtang*, and all connectives in the text are marked and extracted. For instance, the discourse connectives in paragraph (3) are hihglighted and then extracted as in (4). In (4), one line indicates one sentence. The proposition marking punctuation's like ",", ";", ".", "?", etc. are also extracted for showing the proposition boundaries between the connectives. Connectives which are coded with the same first two digits are considered belonging to the same connective-group.

(3) 1.Guan taitai shi ge da gezi, kuai wushi sui le **hal** zhuan ai chuan da hong yifu, **suoyl** waihao jiaozuo dachibaor.
"Madame" Guan was a tall woman. She was almost fifty years old but still loved to wear

2. Chibaor shi ge xiao gua, hongle **yihou**, Beiping de ertong nazhe ta wan.
"Chibaor is a kind of small squash. After it turned red, the children in Beiping liked to play with it.'

3. Zhege waihao qide xiangdang de qiadang, **yinwei** chibaor jing ertong rounong **yihou**, pir **bian** zouqilai, luchu limian de hei zhongzi.
'This nickname was quite appropriate because after being played with by children, the skin of the chibaor became wrinkled, and the inside black seeds were exposed.'

4. Guan Taitai de lianshang **ye** you bu shao de zouwen, **erqie** bizi shang you xuduo queban, **jinguan** ta **hai** chafen mohong, **ye** yanshi bu liao lianshang de zhezi yu heidian.
'Mrs. Guan also had many wrinkles and black spots on her face. No matter how much she powdered and rouged she could not cover up the wrinkles and the black spots.'

5. Ta bi ta de zhangfu de qipai **geng** da, yiju yidong **dou** bo **xiang** Xitaihou.
'Her air was even greater than that of her husband, and each motion and each action was designed to be like the Dowager Empress.'

6. Ta bi Guan xiansheng **geng** xihuan, **ye** geng hui, jiaoji; neng yiqi da liang zheng tian zheng ye de maquepai, **er hai** baochizhe Xitaihou de zunao qidu.
'She liked, even more than Mr. Guan, to cultivate friends and was more capable at this then he. She could at one stretch play mah-jang for two days and two nights and still maintain her loftiness and dignity.' (*Sishi Tongtang* v. 4, p. 18, paragraph 1)

(4) The coding of discourse connectives in paragraph (6.1):

paragraph 1:	sentence 1	,1101hai, 1208suoyi.
	sentence 2	3208yihou,.
	sentence 3	,3304yinwei 3208yihou, 1207bian,.
	sentence 4	1102ye, 4102erqie, 4502jinguan 1101hai, 1102ye.
	sentence 5	1104geng, 1106dou 4301xiang.
	sentence 6	1104geng, 1102ye 1104geng,,1202er 1101hai.

Second, I counted the frequency of each **connective-group** in each **sentence** throughout the entire text. For paragraph (3), as shown in Table 2, in the first sentence, the connective-group 11 (the Emphatic relation in the Additive relations) occurs one time and group 12 (the Consequence relation in the Additive relations) occurs 1 time; group 21 (the Evaluation relation) does not occur; and so on. The frequency of the connective-groups in the other sentences are recorded in the same way.

Table 2 Frequency of Connective-Groups in Sentences 1-6

connective-group coding	11	12	21	31	32	33	41	42	43	44	45	46	47	48	49

sentence	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
	2	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
	3	0	1	0	0	1	1	0	0	0	0	0	0	0	0	0
	4	3	0	0	0	0	0	1	0	0	0	1	0	0	0	0
	5	2	0	0	0	0	0	0	0	1	0	0	0	0	0	0
	6	4	1	0	0	0	0	0	0	0	0	0	0	0	0	0

Similarly, the connective-groups in the paragraph are also counted. The results are listed in (5).

(5) Frequency of Connective-Groups in Paragraph 1:

connective-group coding		11	12	21	31	32	33	41	42	43	44	45	46	47	48	49
paragraph	1	10	3	0	0	2	1	1	0	1	0	1	0	0	0	0

There are 27,549 sentences in total, and 16,010 sentence have connectives. In terms of the scope of paragraphs, there are 6,201 paragraphs in total, and 6,006 paragraphs have connectives.

Fourteen out of 15 connective-groups actually occurred in the text (the exception was group 42, the Generalization relation). Part of them are listed in Table 4 to illustrate the distribution of connective-groups.

Table 4 An Example of Connective-groups in 6,201 Paragraphs

group coding		11	12	21	31	32	33	41	42	43	44	45	46	47	48	49
paragraph	1	10	3	0	0	2	1	1	0	1	0	1	0	0	0	0
	2	4	1	1	0	0	0	1	0	2	0	0	4	0	0	0
	3	3	3	0	0	1	0	1	0	1	0	0	0	1	0	1
	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	6	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
	7	3	1	0	0	0	0	0	0	0	0	0	1	0	0	0
	8	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
	9	0	0	0	2	0	0	0	0	0	0	0	0	0	0	0
	10	1	1	0	0	0	0	0	0	0	0	0	0	1	0	0
	11	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0
	12	1	0	1	0	0	0	1	0	0	0	0	0	0	0	0
	13	2	1	0	0	1	0	0	0	1	0	1	2	0	1	0

Table 3 The Number of Sentences, Paragraphs and the Frequency of Connectives in *Sishi Tongtang*

The number of sentences:	27,549
The number of sentences containing connectives:	16,010
The number of paragraphs:	6,201
The number of paragraphs containing connectives:	6,006
The frequency of connectives:	33,571

As we compare the data in the scope of sentences (as shown in Table 2), and the data in the scope of paragraphs (shown in Table 4), we find them to have one thing in common. Under both scopes, we can see the tendency for some groups of connectives to cooccur with other groups. For instance, group 11 tends to cooccur with group 12 more frequently than with group 21. In addition, in Table 4, the distributions of connective-groups can also show the linear relation between groups; for instance, when group 11 occurs more in a paragraph, group 12 seems to occur more, and when group 11 occurs less, group 12 seems to occur less as well. The distribution of connective-groups in sentences does not reflect this association. Instead, the information about the presence or absence for each connective-group is more prominent under the scope of sentences.

4.1 The Method of Quantifying

Determining the extent to which variation in one variable is related to variation in another is important in many fields of inquiry. Recently the correlation method was applied to linguistic elements for measurements of relatedness in dialect affinity (e.g., Cheng 1973, 1977, 1986). In this study, the numerical measurement of correlation coefficients are used to help us interpret the relations of connectives in coherent texts. I calculate the correlation coefficients between pairs of connectives.

Pearson's correlation coefficient (Glass & Stanley 1970, Kachigan 1986) is appropriate to show the linear relations of the wider range of continuous data. For instance, to calculate the correlation between connectives *suiran* 'although,' and *keshi* 'but' and the correlation between *suiran* 'although' and *yinwei* 'because' based on the frequency of their occurrences in discourse (a) to (e) in (6a), the procedure is illustrated in (6b). The scope of the "discourse unit" here is not specified; it can represent a clause, a sentence-group or any discourse fragment larger in scope. However, units (a) to (e) all represent the same sort of scope.

(6-a)

Discourse	Frequency of the Occurrences *suiran* 'althoug	*keshi'* 'but'	*yinwei* 'because'
a	1	3	0
b	2	4	1
c	0	0	0
d	2	3	2
e	1	1	3

(6-b)

the mean of 'although' = (1+2+0+2+1)/5=1.2
the mean of 'but' = (3+4+0+3+1)/5=2.2
the mean of 'because' = (0+1+0+2+3)/5=1.2

r although-but =

$$\frac{(1-1.2)(3-2.2)+(2-1.2)(4-2.2)+(0-1.2)(0-2.2)+(2-1.2)(3-2.2)+(1-1.2)(1-2.2)}{\sqrt{[(1-1.2)^2+(2-1.2)^2+(0-1.2)^2+(2-1.2)^2+(1-1.2)^2][(3-2.2)^2+(4-2.2)^2+(0-2.2)^2+(3-2.2)^2+(1-2.2)^2]}}$$

$$= 0.8727$$

r although-because =

$$\frac{(1-1.2)(0-12)+(2-1.2)(1-1.2)+(0-1.2)(-1..2)+(2-1.2)(2-1.2)+(1-1.2)(3-1.2)}{\sqrt{[(0-1.2)^2+(1-1.2)^2+(0-1.2)^2+(2-1.2)^2+(3-1.2)^2][(0-1.2)^2+(1-1.2)^2+(0-1.2)^2+(2-1.2)^2+(3-1.2)^2]}}$$

$$= 0.0823$$

As the result shows, the coefficient of 'although' and 'but,' about +0.87, is much higher than the coefficient of 'although' and 'because,' which is about +0.08. The high positive coefficient shows that when 'although' occurs more frequently, 'but' occurs more frequently, and when 'although' occur less frequently, 'but' occurs less frequently. The low positive coefficient between 'although' and 'because,' on the other hand, shows that the occurrences of 'although' are barely associated with the occurrences of 'because.'

The *Jaccard similarity measure*, also known as the *similarity ratio*, was first proposed by

Jaccard in 1908. It has been extensively applied in numerical taxonomies, especially in the field of ecology and bacteriology (Sneath 1973). In lexicostatistics, the *Jaccard similarity measure* has been employed to measure language relations such as in Cheng (1986). The index of Jaccard is related to the task of determining the **presence** or **absence** of a relationship between two random variables. A contingency table of the occurrences of two variants can be constructed to illustrate the correlation of two variants. For example, to see the presence or absence of occurrence between connectives *suiran* 'although' and *keshi*' 'but' in one clause, there could be four possibilities:

- the presence of *suiran* and the presence of *keshi* (+,+)
- the presence of *suiran* and the absence of *suiran* (+,-)
- the absence of *suiran* and the presence of *keshi* (-,+)
- the absence of *suiran* and the absence of *keshi* (-,-)

The above four possibilities can be shown in the form of a 2 x 2 tabular arrangement, often referred to as a contingency table, as in the table below. Beginning with the upper left hand cell and moving in a clockwise direction, the four cells of the table correspond to the (+,+), (+,-), (-,-) and (-,+). In this example, the cases where both *suiran*, and *keshi* are present are 40; that means, in 40 discourse units, *suiran* and *keshi* cooccur. Ten cases in which only *suiran* is present; 20 cases in which both are absent; and 15 cased in which *keshi* is present but *suiran* is not.

The correlation of the pair of connectives can be calculated with the *Jaccard's similarity measure*: the cooccurrences of two variants divided by their total occurrences (Gower 1985).[6] As shown in (5.3), S_j shows the proportion of the sum that mutual presence represents. The correlation of *suiran* and *keshi* is calculated as 0.6154.

(8)

$$S_j = \frac{a}{(a+b+c)}$$

he similarity index of *suiran* and *keshi* $= \frac{40}{(40+10+15)} = 0.6154$

The coefficients are considered as degree of connective-cooccurrence. The correlation coefficients have values ranging from zero to +1. Unlike *Person's coefficient*, the interpretation of *Jaccard's index* is straightforward: The larger the value, the closer are the pair of connectives. Two connectives are closer in the sense that they cooccur more often than other connective pairs. In the case of connective-cooccurrence in clauses, a high coefficient value suggests that connectives X, and Y are more likely to cooccur in one clause. If X is used, it is very likely that Y is also used. That is, they are used more frequently in a proposition to enhance the linkage of an utterance. A low coefficient value, on the other hand, suggests that

(7)

		keshi 'but' +	*keshi* 'but' -
suiran 'though'	+	40 (a)	10 (b)
	-	15 (c)	20 (d)

the two connectives are more likely not to occur in the same clause. This indicates that the use of one connective is more independent of the use of the other connective.

To determine which coefficient method is more appropriate in our study of connective cooccurrence, two aspects need to be considered: (i) whether the data are continuous or dichotomous; and (ii) the purpose of the correlation. The data are continuous when they are any whole number. If the data are either 1 or 0 (i.e., present or absent), the data are dichotomous. Notice that in *Pearson's coefficient*, the **frequency** of connective's occurrence is crucial to decide the coefficient's value. For a positively highly correlated pair of connectives, when one connective occurs more frequently in one clause, the other occurs more frequently and when one occurs less frequently, the other occurs less frequently as well. In *Jaccard's index*, the frequency of a connective's occurrence is not as crucial, instead, the **presence** and **absence** of two connectives in the same clause is essential. *Pearson's coefficient* is appropriate to show the linear relations of the wider range of continuous data, while for the absence or presence of two connectives in one record, the *Jaccard similarity measure* is more suitable. The study of the connectives correlation is based on two different discourse scopes: a proposition and a topic continuity. Within these small scopes of discourse, in most cases, if a connective does occur, it occurs only once. Most of the other connectives do not occur at all. Thus, although the distribution of connectives is based on the frequency of their occurrences, it shows the presence and absence information (further illustrated in Section 5.1.2). Since the data is either 1 or 0 in most cases, *Pearson's correlation* will not be able to capture the association between two connectives. Instead, the *Jaccard similarity measure* can capture the cooccurring information better.[7] Unlike the study of connectives, the distribution of the **groups of connectives** based on the scope of a paragraph really shows the frequency of their occurrences, in most cases, not just 1 or 0. In this case, using *Pearson's correlation* to calculate the linear association between two connectives is more appropriate.

To sum up, the *Jaccard similarity measure* is considered more appropriate for the study of connective cooccurrence in a discourse unit smaller in scope, such as propositions and topic continuities, based on the fact that they are basically dichotomous data. On the other hand, *Pearson's correlation* is adopted for the study of connective-groups in a larger scope, paragraphs, based on the fact that the data are continuous and linearly related.

4.2 The Correlation of Connective-Groups in Sentences

The similarity index of the connective-groups in sentences in the entire book of *Sishi Tongtang* are calculated and listed in Table 5. The higher the coefficient, the closer are the pair

of connective-groups. Two connective-groups are closer than other pairs of connective-groups in the way that they cooccur in a sentence more frequently than the other pairs. For instance, connective-group 11 has a coefficient of 0.221 with group 12, 0.028 with 31, and so on.

The highest 10 rankings of the pairs are listed in Table 6. One thing that needs to be pointed out is that the sequence of a pair of connective-groups is not considered in this data processing. For instance, in the pair of group 11 and group 12, the occurrence of a connective which belongs to group 11 can be either preceded or followed by the group 12 connective; once they cooccur in the same sentence, it counts. However, the sequence of connective-groups in a discourse unit is found to be crucial in their modification directions. It will be further discussed in Section 5.4..

Table 5 Correlation Coefficients of Connective-groups in Sentences in *Sishi Tongtang* (by *Jaccard's Similarity Measures*)

	11	12	21	31	32	33	41	43	44	45	46	47	48
12	.221												
21	.028	.028											
31	.023	.030	.009										
32	.103	.104	.020	.021									
33	.034	.047	.020	.027	.036								
41	.053	.048	.023	.014	.033	.030							
43	.093	.103	.013	.013	.064	.021	.026						
44	.006	.006	.003	.000	.004	.015	.003	.004					
45	.047	.035	.013	.014	.025	.028	.017	.032	.016				
46	.140	.111	.033	.020	.065	.035	.027	.077	.006	.098			
47	.063	.087	.026	.017	.041	.028	.024	.039	.011	.025	.048		
48	.010	.018	.004	.005	.010	.016	.006	.015	.018	.008	.008	.006	
49	.033	.029	.018	.020	.035	.020	.018	.021	.003	.031	.027	.029	.001

4.3 The Correlation of Connective-Groups in Paragraphs

The derived correlation coefficients in the scope of paragraphs are given in Table 7. Although *Pearson's coefficient* ranges from positive 1 to negative 1, in our results, all the coefficients are greater than 0. The positive coefficients indicate that two connective-groups are positively related; namely, when one occurs more in a paragraph, the other occurs more; when one occurs less, the other occurs less. The higher the positive value, the stronger the pair of connective-groups are associated to each other. Table 8 shows the 10 highest ranking pairs of connective-groups.

Table 6 The Highest Ten Ranking of the Correlation Coefficients of Connective-Groups in Sentences

ranking	pair of connective-groups	coefficient value
1	11--12	0.221283
2	11--46	0.140237
3	12--46	0.111423
4	12--32	0.104376
5	11--32	0.103439
6	12--43	0.102879
7	11--43	0.09284
8	45--46	0.098345
9	12--47	0.086678
10	43--46	0.077381

1	11 emphatic -- 12 consequence	(e.g. *ye..jiu* also...then)
2	46 contrastive -- 11 emphatic	(e.g. *keshi..ye* but...also)
3	46 contrastive -- 12 consequence	(e.g. *keshi..jiu* but...and then)
4	32 time -- 12 consequence	(e.g. *shihou..jiu* when...so)
5	32 time -- 11 emphatic	(e.g. *shihou..ye* when...also)
6	12 consequence -- 43 exemplification	(e.g. *jiu..xiang* then...as if)
7	11 emphatic -- 43 exemplification	(e.g. *ye..xiang* also...or example)
8	45 yielding -- 46 contrastive	(eg.*suiran...keshi* although...but)
9	47 general-condition -- 12 consequence	(e.g. *jiaru..jiu* 'if...then')
10	46 contrast -- 43 exemplification	(e.g. *danshi..xiang* but...for example)

5. IMPLICATIONS

In the study of the correlation of connective-groups, all the connectives which denote the same inferential relation are grouped together. To count the correlation of these connective-groups is then to count the correlation of inferential relations in discourse. Thus, a larger picture of the interaction between inferential relations which are marked by the use of connectives, and interpreted by the language user, becomes explicit.

5.1 Sentence vs. Paragraph

The correlations of connective-groups in sentences and in paragraphs, as shown above, are quite similar. Although the coefficient values under the scope of paragraphs is greater than the similarity index derived under the scope of sentences due to the different formulas used, the degrees of closeness indicated in the pairs of connective-groups are generally the same. Compare the highest ten ranking coefficients on both sides, regardless of the slight differences in the ordering, eight out of ten are identical. An implication drawn from this similarity is that discourse connectives as a linkage device are consistently applied by the writer to construct a coherent text no matter whether the text is a sentence long or as long as a paragraph. A paragraph is simply a "larger size" sentence; and the sentence is the smallest unit of a coherent text.

Table 7 Correlation Coefficients of Connective-Groups in Paragraphs in *Sishi Tongtang* (by *Pearson's Correlation Coefficients*)

	11	12	21	31	32	33	41	43	44	45	46	47	48
12	.559												
21	.197	.218											
31	.171	.222	.083										
32	.422	.465	.135	.183									
33	.258	.322	.128	.122	.205								
41	.295	.325	.122	.065	.171	.148							
43	.328	.319	.048	.073	.242	.111	.086						
44	.120	.127	.020	.047	.099	.065	.057	.034					
45	.280	.281	.072	.106	.181	.136	.120	.127	.087				
46	.448	.437	.202	.147	.308	.194	.211	.248	.114	.327			
47	.280	.360	.168	.118	.220	.158	.154	.122	.055	.146	.262		
48	.183	.208	.040	.068	.099	.088	.040	.109	.083	.096	.117	.062	
49	.245	.267	.097	.086	.213	.147	.157	.117	.046	.155	.196	.143	.058

Table 8 The Highest Ten Ranking of the Correlation Coefficients of Connective-Groups in Paragraphs

ranking	pair of connective-groups	coefficient value
1	11--12	0.558516
2	12--32	0.465162
3	11--46	0.447638
4	12--46	0.436788
5	11--32	0.421536
6	12--47	0.359919
7	11--43	0.327704
8	45--56	0.327052
9	12--41	0.324987
10	12--33	0.322232

1	11 emphatic -- 12 consequence	(e.g., *ye..jiu* also...then)
2	32 time -- 12 consequence	(e.g., *shihou..jiu* when...so)
3	11 emphatic -- 46 contrastive	(e.g., *ye..keshi* also...but)
4	12 consequence -- 46 contrastive	(e.g., *jiu..keshi* so...but)
5	32 time -- 11 emphatic	(e.g., *shihou..ye* when...also)
6	12 consequence -- 47 general-condition	(e.g., *jiaru..jiu* if...then)
7	11 emphatic -- 43 exemplification	(e.g., *ye..xiang* also...for example)
8	45 yielding -- 46 contrastive	(e.g., *suiran..keshi* although...but)
9	41 elaboration -- 12 consequence	(e.g., *erqie..jiu* moreover...then)
10	33 cause --12 consequence	(e.g., *yinwei..suoyi* because...so)

BEST COPY AVAILABLE

As discussed in Kuo (1992), in Chinese, there are other pieces of evidence to show that it is the "sentence," not the "paragraph," which is the smallest unit of discourse developing a central topic. The study of correlations in sentences and paragraphs further supports this hypothesis.

5.2 Pedagogical Implication

In addition to the implication discussed above, the correlation values of connective-groups can be used for other purposes. First, concerning language teaching, the ranking of the coefficients provides us with a prioritized list for textbook and material arrangement. In language teaching, connective words are considered essential vocabularies for language learners because they represent the logical linkages between utterances. From the distribution of connectives, readers can pick up the logical flow in discourse easily. And the most efficient way to learn a connective is to learn what other words or patterns this connective usually goes with. For each connective group, the coefficients show the specific degree of closeness with other groups. For instance, to learn how to use contrast connectives, one may want to know how they are used in various situations. From the coefficients index (Table 5), repeated below, we can see that the contrast connectives (46) have higher coefficients with emphatic (11) (with the value of 0.14), consequence (12), (0.111), and yielding connectives (45), (0.098) than other groups. Thus, it may be important to arrange the text material according to the prioritized list.

To teach a particular connective, for instance, *keshi* 'but', teachers can arrange materials according to the prioritized list derived by the correlation of *keshi* 'but' (code 4602) with other connectives as discussed in Chapter 5. For illustration, *keshi*'s highest 20 correlation companions are listed below. For instance, with 1102 *ye* 'and also' the correlation is 0.0232. Teachers can also go further into the running text to show the exact use of *keshi* in the real discourse.

5.3 Reconfirming the Taxonomy of Coherence Relations

Another significance of the coefficients is to reconfirm our taxonomy of coherence relations. Recall that in our theoretical framework, the first task in a successful communication is that "the message itself must be conveyed" and that the Additive relation is used to achieve this task. According to our linguistic knowledge, the Additive relation includes two major logical relations: the Emphatic relation, and the Consequence relation. As the results show, group 11's (additive—emphatic) closest companion is group 12 (additive—consequence) with a coefficient of 0.221. The comparatively high coefficient value of the Emphatic and the Consequence groups reconfirms this taxonomy. Actually, the pair of "emphatic" and "consequence" also has the most frequent occurrences among all the other logical pairs. This suggests that to convey the message itself is actually the most essential step in communication, especially in a narrative

Table 9 The Coefficients of the Contrast Group (46) and th eOther Groups in *Sishi Tongtang*

	em-phatic	conse-quence	evalu-ation	sequence	time	cause	elabo-ration	exempli-fication	alter-nation	yield-ing	genera l-c.	only -c.	all-c.
	11	12	21	31	32	33	41	43	44	45	47	48	49
46	.140	.111	.033	.020	.065	.035	.027	.077	.006	.098	.048	.008	.027

text.

5.4 The Modification Direction of The Inferential Relations

Furthermore, from the distribution of pairs of connectives, it is shown that the sequence of the connectives is crucial. Each inferential relation holds between two adjacent discourse fragments. The discourse fragments may consist of more than one proposition. When an inferential relation holds between two adjacent discourse fragments, the sequences of these two fragments are not always flexible. Li (1990) classifies 116 common-used *quanlian ci* 'relations word' in terms of their syntactic positions into four types: Type A *quanlian ci*'s can only occur in the first clause; Type B can only occur in the second clause; Type C can repeatedly occur in different clauses; and Type D can only occur between two clauses. The examples

Table 10 The Highest 20 Coefficients of Connectives *Keshi* 'but' (Coded as 4602) and the Other Connectives in *Sishi Tongtang*

	1102 *ye*	1101 *hai*	1106 *dou*	1103 *you*	1207 *bian*	4301 *xiang*	1206 *name*	3203 *jinlai*	3304 *yinwei*	3203 *xianzai*
4602	.0232	.0223	.0168	.0128	.0103	.0093	.0084	.0071	.0065	.0064

	1202 *er*	4904 *zong*	1104 *geng*	4304 *fangfu*	3221 *shihou*	3107 *yue*	4112 *ji*	1201 *jiu*	1209 *yinci*	4307 *side*
4602	.0062	.0059	.0059	.0050	.0049	.0047	.0046	.0038	.0035	.0033

is provided below in order to illustrate the four types:[8] (Li, 1990:356)

(9) Type A: Ta **budan** hui Yingwen, ye hui Fawen.
he not-only know English but-also know French
He knows **not only** English, but also French.'

Type B: Wo renshi ta, **danshi** bu da shou.
I know him but not very familiar
'I know him, **but** not very well.'

Type C: **Yaome** ni qu, **yaome** wo qu, kuai jueding.
either you go or I go quickly decide
'**Either** you go **or** I go; make up your mind quickly.'

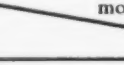
Type D: Zuotian wo jin cheng mai le ji ben shu, **lingwai**, hai qu kan le yi wei pengyou.
yesterday I enter city buy P some C book besides also go see P one C friend
'I went to the city yesterday to buy some books; **besides**, I also visited a friend.'

In this study, I emphasize the directions of modification of each connective-group instead of the syntactic position of each single connective. Each group of connectives involves certain directions in modifying the other discourse fragments. I will call this phenomenon **the prin-**

ciple of Adjacency and further illustrate it below.

For some groups, the discourse chunk marked by the connectives tends to modify only its preceding discourse fragment. For some other groups, the inferential relation holds between the discourse fragment in which the connective occurs and the one following it. For some other groups, the discourse fragment either preceding or following the one marked by the connective can be related to. The modification directions of each inferential relation are illustrated in Table 11. A and B both represent a discourse fragment. Discourse fragment A includes the discourse chunks of different lengths. These can be as small as a proposition, or as large as a complex topic continuity. The discourse connective occurs in either A or B. R represents the inferential relation marked by such a connective. The directions of modification between A and B can be presented in two ways: (i) the fragment containing the connective modifies its preceding fragment, or (ii) the fragment containing the connective modifies its following fragment. When the inferential relation of Emphatic (11), Consequence (12), Sequence (31), Exemplification (43), Alternation(44), or Contrast (46) holds between two discourse fragments,

Table 11 The Modification Directions of Inferential Relations

	inferential relation / direction of modification		
11	B emphasizes A		yes
12	B is the consequence of A		yes
21	A (B) is the evaluation (or comment) of B (A)	yes	yes *
31	A indicates the sequence of information to B		yes
32	A indicates the time information to B	yes	
33	A (B) is the cause of B (A)	yes *	yes
41	B is the elaboration of A	yes	yes *
42	A (B) is the generalization of B (A)	yes *	yes
43	B is the exemplification of A		yes
44	A is the alternation of B		yes
45	A (B) is yielding to B (A)	yes *	yes
46	B is in contrast to A		yes
47	A (B) is the general-condition of B (A)	yes *	yes
48	A (B) is the only-condition of B (A)	yes *	yes
49	A (B) is the all-condition of B (A)	yes *	yes

* : this case occurs more frequently

the one which is marked by the connective is preceded by the one which is modified. For the relations of Time (32), the discourse fragment modifies its following fragment. For other relations, both directions are possible. However, one of the modification directions is more frequent than the other.

In general, discourse connectives have two functions in discourse: the **transition-marking** function and the **inference-marking** function. On the one hand, they are used to mark the connection between the previous and the coming messages and at the same time to introduce the new message to the reader; this is their transition-marking function. The purpose of the connective-groups' modification directions is to provide us with a general picture of the direction of transition-marking. Based on it, the connection between the discourse fragment marked by discourse connectives and its preceding or following discourse can be predicted.

Besides the transition-marking function, on the other hand, discourse connectives are used to mark the particular inference procedure and guide the reader's inference toward a better understanding of the previous message; this is their inference-marking function. For some discourse connectives, the transition-marking function is more apparent than their inference-marking function; for other connectives, it is the other way around; and for some connectives, both ways may occur. When a connective is used to mark the transition function and when it marks the inference function is not crystal clear. Their functions can only be roughly reflected in the taxonomy of inferential relations noted in our previous discussion.

6. CONCLUSION

The numerical measurement of correlation coefficients can be used for different linguistic purposes. In this study, I use the correlation of connective-groups in sentences and in paragraphs to demonstrate four points. First, the similarity between two sets of results reconfirms the hypothesis that in Chinese, the complex sentence represents a topic continuity. Second, the correlation is useful for language teaching purposes. Third, the correlation result reconfirms our taxonomy of coherent relations. And fourth, and most importantly, from the distribution of pairs of connective-groups, the modification direction for the inferential function denoted by each connective-group can be generalized. This generalization, the Adjacency principle, tells us the direction of the scope covered by discourse connectives. It will be the base for establishing the discourse structure in terms of logical linkages.

ACKNOWLEDGEMENTS

I wish to thank C. C. Cheng, Yamura Kachru, Fumiyoshi Matsumura, and Wen-Chiu Tu for their valuable comments on the earlier versions of this paper, and to Fred Davidson for his suggestions concerning the statistical aspect of my research design. The responsibility for any errors and omissions rests with me.

THE AUTHOR

Pinmin Kuo is a graduate student in Linguistics of the University of Illiinois at Urbana-Champaign.

NOTES

[1] This term is defined by Chao (1968). Unlike intrasentential syntactic conjunctions, macrosyntactic conjunctions function intersententially.

[2] Abbreviations in the glosses: P = particle, Nom = nominalizer, C = classifier, Q = question marker.

[3] Examples are taken from Lao She's (1982) *Luotuo Xiangzi*.

[4] The text database of *Sishi Tongtang* was created by Fumiyoshi Matsumura. For the details of the creation see Matsumura (1992). However, I am wholly responsible for the indexing process and the data application.

[5] In this paper, a sentence (or a sentence-group, 'ju qun' in Chinese) represents a basic topic continuity. And paragraph 'duanluo' represents a complex topic continuity. See Kuo (1994) for more discussion.

[6] The cases that both variants are absent (-,-) are excluded in *Jaccard's similarity measure*. In her study of dialect classification, Tu (1994) compares *Jaccard's similarity measure* with phi coefficients and *Ellegard's correlation* based on the quantitative method discussed in Cheng (1986). In her discussion, *Jaccard's similarity measure* is preferred over phi coefficients and *Ellegard's correlation* based on the facts that the former "excludes (0,0), does not derive infinity, and treats (+,+), (+,-) and (-,+) equally" (Tu 1994). In this study, phi coefficients and *Ellegard's correlation* are not considered based on this same reason.

[7] In the calculation, when the frequency of occurrences is 1 or greater than 1, the present index '1' is marked; when no connective occurs, the absent index '0' is given.

[8] Li (1990) is in Chinese. The translation of these example is mine.

REFERENCES

Chao, Y. (1968). *A grammar of spoken Chinese*. University of California Press.

Chen, H. (Ed). (1989). *Zhongguo Yuyan Xue Da Cidian (Encyclopedic Dictionary of Chinese Linguistics)*. Jiangxi Jiaoyu ChubanShe.

Cheng, C. (1986). *Quantifying affinity among Chinese dialects. Technical Report No. LLL-T-10-86*. Language Learning Laboratory, University of Illinois at Urbana-Champaign.

_____. (1986). *Quantifying affinity among Chinese dialects. Technical Report No. LLL-T-10-86*. Language Learning Laboratory, University of Illinois at Urbana-Champaign

Ding, B. (1980). *Zhongguohua de Wenfa. (The Chinese Translation of Yuen-Ren Chao's A Grammar of Spoken Chinese*. University of California Press). Hong Kong: Chinese University Press.

Glass, G. V., and Stanley, J.C. (1970). *Statistical Methods in Education and Psychology*. Englewood Cliffs, New Jersey: Prentice-Hall, Inc.

Gower, J.C. (1985). Measures of similarity, dissimilarity, and distance. In S. Kotz, N. Lloyd and C.B. Read (Eds.), *Encyclopedia of Statistical Sciences: Volume 5: Lindberg Condition to Multitrait-Multimethod Matrices* (pp. 397-405). New York: John Wiley and Sons.

Guo, Y. (1960). *Fuci, Jieci, Lianci (Adverbs, Prepositions, and Conjunctions)*. Shanghai Jiaoyu Chubanshe.

Hobbs, J. (1978). *Coherence & coreference. SRI Technical Note 168, SRI International.* Menlo Park, California, August 1978.

_____. (1979). Why is discourse coherent? In Neubauer (Ed.), *Coherence in Natural Language Texts.* Gesamtherstellung: WS Driclereo Werner Schaubruch, Mainz.

James, J. (1979). *Rickshaw.* Translation of Lao She's novel Luotuo Xiangzi. University Press of Hawaii.

Jing, S. (1980). *Xiangdai Hanyu Xuci (Modern Mandarin Chinese Empty Words).* Naimengu: Renmin Chu Ban She.

Kachigan. (1986). *Statistical Analysis.* Radius Press.

King, E. (1945). *Rickshaw Boy. translation of Lao She's Luotuo Xiangzi 1935.* New York: Reynal & Hitchcock.

Kuo, P. (1992). *Sentence, paragraph and topic continuity: a contrastive study of discourse structure in Mandarin Chinese.* Paper presented in the 25th International Conference on Sino-Tibetan Languages and Linguistics. University of California Berkeley.

_____. (1994). *Discourse Connectives and Discourse Structure in Mandarin Chinese.* University of Illinois Doctoral Dissertation, in press.

Lao S. (1982). Luotuo Xiangzi in Laoshe Xuangji vol. 1, pp. 1-243. (first print in Yu Zhou Feng 1936-1937) Chengdu: Sichuan Ren Min Chuban She.

_____. (1983). Huang Kong (Fear) in Lao She Wenji (The Selection of Lao She) vol. 4 (first printed in Sao Dang Bao, 1946) Beijing: Renmin Wenxue Chubanshe.

_____. (1983). Tou Sheng (Coward) in Lao She Wenji (The Selection of Lao She) vol. 5 (first printed in Shi Jie Ribao, 1946) Beijing: Renmin Wenxue Chubanshe.

_____. (1984). Ji Huang (Famine) in Lao She Wenji (The Selection of Lao She) vol. 6 (first printed in Xiao Shuo, 1950) Beijing: Renmin Wenxue Chubanshe.

Lee, C. (1990). Recovery & Translation of Zero Anaphoric Subjects in Chinese. University of Illinois Doctoral Dissertation.

Li, C. N. & Thompson, S. A. (1981). *Mandarin Chinese: A functional reference grammar.* University of California Press.

Li, X. (1990). Xiandai Hanyu fuju zhong guanlian ci de weizhi (The positions of relation words in complex sentences in modern Chinese). DiSan Jie Guoji Hanyu Jiaoxue Taolunhui Lunwenxuan. Beijing Hanyu Xueyuan Chubanshe.

Lü, S. (1980). *Xiandai Hanyu Ba Bai Ci (Eight Hundred Words in Modern Chinese)* Shangwu Yingshu Guan Chuban.

Matsumura, F. (1992, October). *From a noun to verbs and from a verb to nouns.* Paper presented in the 25th International Conference on Sino-Tibetan Languages and Linguistics University of California, Berkeley.

Okurowshi, M. E. (1986). *Textual cohesion in Modern Standard Chinese.* Georgetown University Doctoral Dissertation.

Schiffrin, D. (1987). *Discourse marker.* Cambridge: Cambridge University Press.

Schourup, L. C. (1985). *Common discourse particles in English conversation.* New York: Garland Publishing.

Shi, X. (1981). *Camel Xiangzi (Translation of Laoshe's Luotuo Xiangzi 1935).* Indiana University Press & Foreign Languages Press.

Sneath, P. H. A. & Sokal, R. R. (1973). *Numerical Taxonomy.* San Francisco: W. H. Freeman and Company.

Tu, W. (1994). *Rukai Mutual Intelligibility and Classification.* University of Illinois Doctoral Dissertation, in press.